UNIVERSIDAD
DE GRANADA

Escuela Técnica Superior de Ingenierías Informática y de Telecomunicación
Facultad de Ciencias

Doble Grado en Ingeniería Informática y Matemáticas

trabajo de fin de grado

Optimización de redes neuronales

Presentado por:
Blanca Cano Camarero

Tutores:
Juan Julián Merelo Guervós
Arquitectura y tecnología de computadores

Francisco Javier Meri de la Maza
Análisis matemático

Curso académico 2021-2022

Optimización de redes neuronales

Blanca Cano Camarero

Blanca Cano Camarero *Optimización de redes neuronales*.
Trabajo de fin de Grado. Curso académico 2021-2022.

Responsables de tutorización

Juan Julián Merelo Guervós
Arquitectura y tecnología de computadores

Francisco Javier Meri de la Maza
Análisis matemático

Doble Grado en Ingeniería Informática y Matemáticas

Escuela Técnica Superior
de Ingenierías Informática
y de Telecomunicación
Facultad de Ciencias

Universidad de Granada

Índice general

Índice de figuras

Índice de tablas

Parte I.

Teoría subyacente

1. Teoría de la aproximación

1.1. Objetivos

El desarrollo de los capítulos comprendidos entre Polinomios de Bernstein 1.3, a la demostración del teorema de Stone Weierstass 1.4 es múltiple. Se pretende primeramente construir las herramientas esenciales para la demostración del Teorema Universal de redes neuronales por propagación hacia delante y hacia detrás; mas comprendiendo la naturaleza del fundamento es posible entender la bondad, alcance e imposición de las estructuras elementales que conforman las redes neurales, luego se hará simultaneamente un análisis y estudio de las implicaciones de la teoría demostrada.

1.2. Bibliografía

La documentación consultada para esta sección ha sido:

1. Para demostraciones básicas [**?**].

1.3. Polinomios de Bernstein

En esta seción introduciremos los polinomios de Bernstein; que vistos como una serie nos asegurarán una convergencia uniformemente a cualquier función continua en un compacto y serán esenciales para nuestra prueba del teorema de Stone-Weiertrass.

Comenzaremos recordando el Teorema del Binomio de Newton.

Teorema 1.1 (Binomio de Newton)**.** *Cualquier potencia de un binomio $x + y$ con $x, y \in \mathbb{R}$, puede ser expandido en una suma de la forma*

$$(x+y)^n = \sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k}$$

Tomando ahora para esta igualdad $x \in \mathbb{R}, y = 1 - x$ se tiene que

$$1 = (x + (1-x))^n = \sum_{k=0}^{n} \binom{n}{k} x^k (1-x)^{n-k} \tag{1.1}$$

Dada cualquier función f definida en x podríamos multiplicar la ecuación (1.1) por $f(x)$ resultando.

$$f(x) = \sum_{k=0}^{n} f(x) \binom{n}{k} x^k (1-x)^{n-k} \tag{1.2}$$

Y tomando como dominio $I = [0,1]$ de $f : I \longrightarrow \mathbb{R}$, nos encontramos frente a una ecuación muy sugerente para introducir $B_n(x)$, el *Polinomio n-ésimo de Bernstein* . El cual pretende aproximar la función f a través de los puntos $\frac{k}{n}$ con $n \in \mathbb{N}$ fijo y $k \in \{0, ..., n\}$.

Definición 1.1 (Polinomios de Bernstein)**.** Dada cierta función $f : [0,1] \to \mathbb{R}$, se define el n-ésimo polinomio de Bernstein para f como

$$B_n(x) = B_n(x; f) = \sum_{k=0}^{n} \left(f\left(\frac{k}{n}\right) \binom{n}{k} x^k (1-x)^{n-k} \right).$$

Faltaría por ver si efectivamente nuestro polinomio construido *aproxima lo suficientemente bien* a la f originaria. Basándonos en la igualdad (1.2) y la diferencia entre $f(x)$ y $B_n(x)$ se concluye que

$$|f(x) - B_n(x)| = \sum_{k=0}^{n} \left| f(x) - f\left(\frac{k}{n}\right) \right| \binom{n}{k} x^k (1-x)^{n-k}$$

Ante tal igualdad la intuición ya nos hace pensar que sea convergente al aumenta el tamaño de la partición. En efecto, en nuestro sucesivo teorema, nos cercioraremos que 1.1 es uniformemente convergente a f en un compacto.

Teorema 1.2 (Teorema de aproximación de Bernstein)**.** *Sea f una función continua de un intervalo $I = [0,1]$ con imágenes en los reales. La secuencia de polinomio de Bernstein 1.1 converge uniformemente a f en I.*

Recordaremos antes la definición de convergencia uniforme:

Definición 1.2 (Convergencia uniforme para funciones reales)**.** Dado E un conjunto y $\{f_n\}_{n\in\mathbb{N}}$ una sucesión de funciones de E a los reales; se dice que dicha sucesión converge uniformemente si para cualquier $\varepsilon > 0$ existe un número natural m tal que para todo $x \in E$ y cualquier natural n que cumpla $n \geq m$ se tiene que

$$|f_n(x) - f(x)| < \varepsilon$$

Comencemos pues con la demostración del teorema 1.2.

Demostración. Para cualquier $\varepsilon > 0$ queremos probar que existe un $m_\varepsilon \in \mathbb{N}$ tal que para todo $x \in I$ e $n \geq m_\varepsilon$ se tiene que $|f(x) - B_n(x)| < \varepsilon$.

Para ello por estar f definida en un intervalo, se tienen dos consecuencias claves:

1. Está acotada, supongamos por $M \in \mathbb{R}$, esto es $|f(x)| \leq M$.
2. En virtud del teorema de Heine-Cantor f es uniformemente continua, es decir; por estar f definida en un compacto, para cualquier $\varepsilon > 0$ existirá un δ_ε tal que para cualesquiera $x, y \in I$ que cumplan $|x - y| < \delta_\varepsilon$ entonces $|f(x) - f(y)| < \varepsilon$.

En virtud de la consecuencia 1. Dado $N \in \mathbb{N}$ fijo pero arbitrario, para cualquier $k \in \{1, ..., N\}$ se tiene que $\frac{k}{N} \in I$ y tomando $x \in I$ podemos acotar por la desigualdad triangular

$$\left| f(x) - f\left(\frac{k}{N}\right) \right| \leq |f(x)| + \left| f\left(\frac{k}{N}\right) \right| \leq 2M$$

Por lo que

$$|f(x) - B_n(x)| \sum_{k=0}^{n} \left| f(x) - f\left(\frac{k}{n}\right) \right| \leq \binom{n}{k} x^k (1-x)^{n-k}$$

Puesto que tenemos que f acotada por M y es uniformemente continua, para valores de k tales que $\frac{k}{n}$ esté próxima a x, tal término de la sumatoria será pequeño por la continuidad de f en x. Por otro lado si está lo suficientemente alejado, tan solo podremos acotar tal término por $2M$.

Separaremos pues nuestra sumatoria en los siguientes dos conjuntos.

Para $\varepsilon > 0$ y para δ_ε de la definición de continuidad uniforme de f podemos encontrar un $n \geq \sup\{(\delta_\varepsilon)^{-4}, \frac{M^2}{\varepsilon^2}\}$. TODO: explicar porqué se eleva a menos cuatro, tiene que ver con la cota de los mayores.

Dispuestos a separar la sumatoria en en función de la distancia mencionada resultan los conjuntos:

$$\mathcal{A}_{n,x} = \{k \text{ tales que } k \in \{0, ..., n\} \text{ y } |x - \frac{k}{n}| < n^{\frac{-1}{4}} \leq \delta_\varepsilon$$
$$\mathcal{B}_{n,x} = \{0, ..., n\} - \mathcal{A}$$

.

De donde obtenemos la siguiente estimación:

$$\sum_{k\in\mathcal{A}} \left| f(x) - f\left(\frac{k}{n}\right) \right| \binom{n}{k} x^k (1-x)^{n-k} \leq \sum_{k\in\mathcal{A}} \varepsilon \binom{n}{k} x^k (1-x)^{n-k} =$$
$$= \sum_{k\in\mathcal{A}} \varepsilon \binom{n}{k} x^k (1-x)^{n-k} \leq \varepsilon \sum_{k=0}^{n} \binom{n}{k} x^k (1-x)^{n-k} = \varepsilon$$

Para el resto de sumandos para los que $|x - \frac{k}{n}| \geq n^{\frac{-1}{4}}$ se tiene que $(x - \frac{k}{n})^2 \geq n^{\frac{-1}{2}}$ cota que utilizaremos más adelante.

$$|f(x) - B_n(x)| \sum_{k=0}^{n} \left| f(x) - f\left(\frac{k}{n}\right) \right| \leq \binom{n}{k} x^k (1-x)^{n-k} \leq \sum_{k=0}^{n} 2M \binom{n}{k} x^k (1-x)^{n-k}$$

Encauzados nuevamente por la intuición de que la diferencia debe decrecer al aumentar N, el tamaño de la partición; escribiremos la ecuación anterior en involucrando a N.

El siente paso natural será acotar la sumatoria entera por alguna expresión que decrezca al aumentar N, para ello jugaremos un poco con la propiedades de los coeficientes binomiales. TODO : es necesario introducir aquí la inecuaciónde la N, pero ahora no ves clara la relación, conviene y quizás de aquí surgen la necesidad de la partición.

Tengamos ahora presente las siguientes igualdades

$$\binom{n-1}{k-1} = \frac{(n-1)!}{(k-1)!(n-1-(k-1))!} = \frac{k}{n}\binom{n}{k} \tag{1.3}$$

$$\binom{n-2}{k-2} = \frac{(n-2)!}{(k-2)!(n-2-(k-2))!} = \frac{k(k-1)}{n(n-1)}\binom{n}{k} \tag{1.4}$$

Partiendo de la igualdad (1.1):

$$1 = (x + (1-x))^n = \sum_{k=0}^{n} \binom{n}{k} x^k (1-x)^{n-k}$$

Reemplazamos la n por $n-1$ y la k por j y tenemos

$$1 = \sum_{j=0}^{n-1} \binom{n-1}{j} x^j (1-x)^{(n-1)-j}$$

Multiplicamos por x y aplicamos la igualdad (1.3) resultando

$$x = \sum_{j=0}^{n-1} \frac{j+1}{n} \binom{n}{j+1} x^{j+1} (1-x)^{(n-(j+1)}$$

Renombramos $k = j+1$, por lo que resulta

$$x = \sum_{k=1}^{n} \frac{k}{n} \binom{n}{k} x^k (1-x)^{n-k}$$

Como el término con $k = 0$ es nulo podemos añadirlo a la sumatoria

$$x = \sum_{k=0}^{n} \frac{k}{n} \binom{n}{k} x^k (1-x)^{n-k} \tag{1.5}$$

Haremos ahora un razonamiento similar sustituyendo n por $n-2$

Partiendo de (1.1) se tiene que

$$1 = (x + (1-x))^n = \sum_{k=0}^{n} \binom{n}{k} x^k (1-x)^{n-k}$$

Reemplazamos la n por $n-2$ y la k por j y tenemos

$$1 = \sum_{j=0}^{n-2} \binom{n-2}{j} x^j (1-x)^{(n-2)-j}$$

Multiplicamos por x^2 y aplicamos la igualdad (1.4) resultando

$$x^2 = \sum_{j=0}^{n-2} \frac{(j+2)(j+1)}{n(n-1)} \binom{n}{j+2} x^{j+2} (1-x)^{(n-(j+2))}$$

Renombramos $k = j+2$, por lo que resulta

$$x^2 = \sum_{k=2}^{n} \frac{k(k-1)}{n(n-1)} \binom{n}{k} x^k (1-x)^{n-k}$$

Como con los términos $k=0$ y $k=1$ se anula, podemos añadir dichos índices sin modificar la suma

$$x^2 = \sum_{k=0}^{n} \frac{k(k-1)}{n(n-1)} \binom{n}{k} x^k (1-x)^{n-k}$$

Podemos reescribir la ecuación resultando:

$$(n^2-n)x^2 = \sum_{k=0}^{n} (k^2-k) \binom{n}{k} x^k (1-x)^{n-k} \tag{1.6}$$

$$\begin{aligned}
&\sum_{k\in\mathcal{B}_{x,N}} 2M \binom{N}{k} x^k (1-x)^{N-k} \\
&= 2M \sum_{k\in\mathcal{B}_{x,N}} \frac{(x-\frac{k}{N})^2}{(x-\frac{k}{N})^2} \binom{N}{k} x^k (1-x)^{N-k} \\
&\leq 2M\sqrt{N} \sum_{k\in\mathcal{B}_{x,N}} (x-\frac{k}{N})^2 \binom{n}{k} x^k (1-x)^{N-k}
\end{aligned}$$

Recordemos que nuestro objetivo era acotar 1.3
Para ello vamos a sumar las dos expresiones que hemos obtenido (1.5) y (1.6) resultando

$$(n^2-n)x^2 + nx = \sum_{k=0}^{n} ((k^2-k)+k) \binom{n}{k} x^k (1-x)^{n-k}$$

Dividimos todo entre n.

$$(1-\frac{1}{n})x^2 + \frac{1}{n}x = \sum_{k=0}^{n} \left(\frac{k}{n}\right)^2 \binom{n}{k} x^k (1-x)^{n-k} \tag{1.7}$$

A continuación sumamos a la igualdad (1.7) la ecuación (1.1) multiplicada por x^2 y la ecuación (1.5) multiplicada por $-2x$ resultando:

$$(1-\frac{1}{n}+1-2)x^2 + \frac{1}{n}x = \sum_{k=0}^{n} \left(\left(\frac{k}{n}\right)^2 + x^2 - 2x \right) \binom{n}{k} x^k (1-x)^{n-k}$$

Factorizando en ambos miembros resulta

$$\frac{1}{n}x(1-x) = \sum_{k=0}^{n} \left(x - \frac{k}{n}\right)^2 \binom{n}{k} x^k (1-x)^{n-k} \tag{1.8}$$

Gracias a (1.8) acabos de encontrar una cota para (1.3)

$$2M\sqrt{N} \sum_{k \in \mathcal{B}_{x,N}} (x - \frac{k}{N})^2 \binom{N}{k} x^k (1-x)^{N-k} = 2M\sqrt{N}\frac{1}{N}x(1-x) = 2M\frac{1}{\sqrt{N}}x(1-x)$$

Además como $x(1-x)$ alcanza un máximo absoluto en $x = \frac{1}{2}$ luego concluimos que

$$|f(x) - B_N(x)| = \sum_{k=0}^{N} \left| f(x) - f\left(\frac{k}{N}\right) \right| \binom{n}{k} x^k (1-x)^{n-k} \leq M\frac{1}{2\sqrt{N}}$$

Con $M = max\{|f(x)| : x \in I\}$.

Por que tomando como $m_\epsilon = \lceil M\frac{1}{2\sqrt{\epsilon}} \rceil$ habremos probado la convergencia uniforme buscada.

$\square$

Realizando un repaso global habiendo acabado el teorema, se pueden extraer que conjunto a un ingenioso manejo de operaciones y acotaciones; la clave del resultado reside en las consideraciones en 1 2 y estas a su vez en la compacidad de I.

Por su parte, la selección del dominio de $I = [0, 1]$ viene determinada ya que los nodos $\{\frac{k}{N} | k \in \{0, ..., N\} | \}$ sobre los que se construye el *N-ésimo polinomio de Bernstein* deben pertenecer a I.

Sin embargo, tal dificultad es facilmente salvable con un homeomorfismo.

Como resultado de relajar el dominio donde se define f, pidiéndole tan solo compacidad nace el siguiente corolario.

Corolario 1.1 (Teorema de aproximación de Weierstass)**.** *Sea f una función continua definida en un intervalo real y con valores reales. Se tiene que f puede ser aproximada uniformemente con polinomios.*

Demostración. Si f se encuentra definida en $[a, b]$ con $a < b$ y bastará considerar la función

$$g(t) = f((b-a)t + a) \text{ con } t \in [0, 1]$$

Esta función está definida en $[0, 1]$, tiene la misma imagen que f y mantiene la continuidad ya que se ha construido a través del homeomorfismo $H : [0, 1] \longrightarrow [a, b]$, con $H(t) = (b-a)t + a$.

En virtud del teorema de convergencia 1.2 g podrá ser aproximada uniformemente por una sucesión de polinomios de Bernstein $\{S_n\}_{n \in \mathbb{N}}$. Gracias a ella se puede construir la sucesión $\{S_n \circ H^{-1}\}_{n \in \mathbb{N}}$, que aproxima uniformemente a f. $\square$

1.4. Teorema de Stone-Weierstrass

Teorema 1.3 (Teorema de Stone-Weierstrass)**.** *Sea K un subconjunto compacto de $\mathbb{R}^p$ y sea $\mathcal{A}$ una colección de funciones continuas de K a $\mathbb{R}$ con las siguientes propiedades:*

1. *La función constantemente uno, definida como $e(x) = 1$, para cualquier $x \in K$ pertenece a $\mathcal{A}$.*
2. *Cerrado para sumas y producto para escalares reales. Si f, g pertenece a $\mathcal{A}$, entonces $\alpha f + \beta g$ pertenece a $\mathcal{A}$.*
3. *Cerrado para producto. Para $f, g \in \mathcal{A}$, se tiene que fg pertenece a $\mathcal{A}$.*
4. *Separación de K, es decir si $x \neq y$ pertenecientes a K, entonces existe una función f en $\mathcal{A}$ de tal manera que $f(x) \neq f(y)$.*

Se tiene que toda función continua de K a $\mathbb{R}$ puede ser aproximada en K por funciones de $\mathcal{A}$.

La idea que subyace bajo la demostración del teorema es que a partir de las hipótesis de la estructura algebraica y separabilidad se prueba que es posible encontrar las funciones máximo y mínimo, construyéndolo a partir del valor absoluto.

Con esta nueva propiedad y siendo cerrado para combinaciones lineales se puede aproximar uniformemente cualquier función continua definida en el compacto.

Demostración. Sean $a, b \in \mathbb{R}$ y $x \neq y$ pertenecientes a K. Por la hipótesis de separabilidad existirá un función $f \in \mathcal{A}$ tal que $f(x) \neq f(y)$. Además la existencia de un elemento neutro $e \in \mathcal{A}$ en el álgebra nos permite encontrar reales α, β tales que

$$\alpha f(x) + \beta e(x) = a, \quad \alpha f(y) + \beta e(y) = b$$

Por el teorema de Heine, para $f \in \mathcal{A}$ está acotada por tomar imagen en un compacto, es decir $|f(x)| \leq M$ para $x \in K$.

Consideremos ahora la función valor absoluto, $\phi(t) = |t|$ definida en el dominio $I = [-M, M]$. Por el teorema de aproximación de Weierstrass 1.1

para cualquier $\varepsilon > 0$ existirá un polinomio p cumpliendo que

$$||t| - p(t)| < \varepsilon, \quad \forall t \in I.$$

Puesto que $t \in I$ no son más que las posibles imágenes que puede tomar f en K inferimos entonces que

$$||f(x)| - p \circ f(x)| < \varepsilon \quad \forall x \in K.$$

Como $f \in \mathcal{A}$ y p es un polinomio, es decir, $p \circ f(x)$ son sumas de potencias multiplicadas por escalares de $f(x)$, luego por la hipótesis de ser cerrado a estas operaciones tenemos que la función $|f|$ pertenece a $\mathcal{A}$ si $f \in \mathcal{A}$.

Tenemos con esto que $\mathcal{A}$ también es cerrada a supremo e ínfimo gracias a que:

$$sup\{f, g\} = \frac{1}{2}\{f + g + |f + g|\}$$

$$inf\{f, g\} = \frac{1}{2}\{f + g - |f + g|\}$$

Por lo que concluimos que cualquier función puede ser uniformemente aproximada por combinaciones lineales, supremos e ínfimo mediante funciones de $\mathcal{A}$. Es decir, cualquier función continua en K puede ser uniformemente aproximado por funciones de $\mathcal{A}$. $\square$

Admite este teorema una reflexión sobre cuánto de restrictivas son las hipótesis o el conjunto de propiedades exigidas $\mathcal{A}$.

Para ello tomemos un ejemplo concreto, sea $K = \mathbb{R}^n$ se tiene que el conjunto $\mathcal{A} = \{\}$ Reflexión sobre el teorema, K puede parecer todo lo abstracta posible, ¿Son por ende las hipótesis demasiado restrictivas? Para responder a esta pregunta basta con pensar en un ejemplo concreto: Tomemos $K = \mathbb{R}^n$ el cuerpo de los polinomios cumple todas las hipótesis, luego...

Sin embargo aproximar uniformemente cualquier función continua es Lo restrictivo está en pensar ya que el conjunto de funciones continuas en un compacto es mucho más pequeño que el de todas las funciones que podamos pensar.

Agradecimientos

Agradezco a